Attorney Docket No.: 0160115

UNITED STATES PATENT APPLICATION

FOR

ADAPTIVE CORRELATION WINDOW FOR OPEN-LOOP PITCH

INVENTOR:

YANG GAO

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service "Express Mail Post Office to addressee" Service under 37 C.F.R. Sec. 1.10 addressed to: Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450, on 3/11/04

Express Mailing Label No.: EV420421900US

Lori Lapidario — Lori Lapidario

Name — Signature

PREPARED BY:

FARJAMI & FARJAMI LLP
26522 La Alameda Ave., Suite 360
Mission Viejo, California 92691

(949) 282-1000
Customer No. 25700

25700

PATENT TRADEMARK OFFICE

ADAPTIVE CORRELATION WINDOW FOR OPEN-LOOP PITCH

RELATED APPLICATIONS

The present application claims the benefit of United States provisional application serial number 60/455,435, filed March 15, 2003, which is hereby fully incorporated by reference in the present application.

United States Patent Application Serial Number _____, "SIGNAL DECOMPOSITION OF VOICED SPEECH FOR CELP SPEECH CODING," Attorney Docket Number: 0160112.

United States Patent Application Serial Number _____, "VOICING INDEX CONTROLS FOR CELP SPEECH CODING," Attorney Docket Number: 0160113.

United States Patent Application Serial Number _____, "SIMPLE NOISE SUPPRESSION MODEL," Attorney Docket Number: 0160114.

United States Patent Application Serial Number _____, "RECOVERING AN ERASED VOICE FRAME WITH TIME WARPING," Attorney Docket Number: 0160116.

BACKGROUND OF THE INVENTION

1. FIELD OF THE INVENTION

The present invention relates generally to speech coding and, more particularly, to pitch correlation of voiced speech.

2. RELATED ART

From time immemorial, it has been desirable to communicate between a speaker at one point and a listener at another point. Hence, the invention of various

telecommunication systems. The audible range (i.e. frequency) that can be transmitted and faithfully reproduced depends on the medium of transmission and other factors. Generally, a speech signal can be band-limited to about 10 kHz without affecting its perception. However, in telecommunications, the speech signal bandwidth is usually limited much more severely. For instance, the telephone network limits the bandwidth of the speech signal to between 300 Hz to 3400 Hz, which is known in the art as the "narrowband". Such band-limitation results in the characteristic sound of telephone speech. Both the lower limit at 300Hz and the upper limit at 3400 Hz affect the speech quality.

In most digital speech coders, the speech signal is sampled at 8 kHz, resulting in a maximum signal bandwidth of 4 kHz. In practice, however, the signal is usually band-limited to about 3600 Hz at the high-end. At the low-end, the cut-off frequency is usually between 50 Hz and 200 Hz. The narrowband speech signal, which requires a sampling frequency of 8 kb/s, provides a speech quality referred to as toll quality. Although this toll quality is sufficient for telephone communications, for emerging applications such as teleconferencing, multimedia services and high-definition television, an improved quality is necessary.

The communications quality can be improved for such applications by increasing the bandwidth. For example, by increasing the sampling frequency to 16 kHz, a wider bandwidth, ranging from 50 Hz to about 7000 Hz can be accommodated. This bandwidth range is referred to as the "wideband". Extending the lower frequency range to 50 Hz increases naturalness, presence and comfort. At the other end of the spectrum, extending the higher frequency range to 7000 Hz increases intelligibility and makes it easier to differentiate between fricative sounds.

Digitally, speech is synthesized by various well-known methods. One popular

method is the Analysis-By-Synthesis (ABS) method. Analysis-By-Synthesis is also referred to as closed-loop approach or waveform-matching approach. It offers relatively better speech coding quality than other approaches for medium to high bit rates. One ABS approach is the so-called Code Excited Linear Prediction (CELP) method. In CELP coding, speech is synthesized by using encoded excitation information to excite a linear predictive coding (LPC) filter. The output of the LPC filter is compared against the voiced speech and used to adjust the filter parameters in a closed loop sense until the best parameters based upon the least error is found.

Pitch lag is one of the most important parameters for voiced speech, because the perceptual quality is very sensitive to pitch lag. CELP speech coding approaches rely on determination of open-loop pitch to help minimize the weighted errors in the closed-loop speech coding process. Open-loop pitch is usually determined using normalized pitch correlation on a weighted speech signal. With this approach, it is desirable to maximize correlation between a windowed reference signal and a candidate signal. Thus, the correlation window size is traditionally limited to have a good local pitch lag, a reliable determination of small pitch lags, and acceptable complexity. However, because voiced speech is not purely periodic, this approach may fail when the local pitch lag is larger than the window size and/or when an energy peak is not located within the window.

The present invention addresses the issues identified above regarding pitch lag determination.

SUMMARY OF THE INVENTION

In accordance with the purpose of the present invention as broadly described herein, there is provided systems and methods for adaptively adjusting the correlation window for open-loop pitch determination.

Generally, for CELP speech coding, open loop pitch is determined using a normalized pitch correlation approach. In order to minimize weighted errors in the closed-loop process (e.g. CELP coding), pitch lag is estimated on the weighted speech signal. However, sometimes the correlation window for pitch lag estimation may fail to contain a complete pitch cycle thus making correlation difficult. If the window is too large, it may cause complexity problem and also increase the difficulty to detect a short pitch lag. Embodiments of the present invention provide methods to maximize correlation between a windowed reference signal and a candidate signal under most conditions by sliding the window by a delta increment in either direction to capture peak energy. The traditional fixed size of the correlation window is maintained. However, the window slides forward and/or backward to capture peak energy within the window.

In one embodiment of the present invention, the position of the adjusting or sliding window may shift in a small range or increment to maximize the energy of the windowed signal thus making sure that at least one peak energy is captured within the window. The methods of the present invention correct the possible errors in detection of large pitch lags without affecting the reliability of detecting small pitch lags.

These and other aspects of the present invention will become apparent with further reference to the drawings and specification, which follow. It is intended that all such additional systems, methods, features and advantages be included within this description, be within the scope of the present invention, and be protected by the

accompanying claims.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 is an illustration of the windowing of a time domain representation of the energy of a coded voiced speech signal.

Figure 2 is an illustration of the sliding window concept in accordance with an embodiment of the present invention.

Figure 3 is a flowchart illustration of a positive sliding window in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION

The present application may be described herein in terms of functional block components and various processing steps. It should be appreciated that such functional blocks may be realized by any number of hardware components and/or software components configured to perform the specified functions. For example, the present application may employ various integrated circuit components, e.g., memory elements, digital signal processing elements, transmitters, receivers, tone detectors, tone generators, logic elements, and the like, which may carry out a variety of functions under the control of one or more microprocessors or other control devices. Further, it should be noted that the present application may employ any number of conventional techniques for data transmission, signaling, signal processing and conditioning, tone generation and detection and the like. Such general techniques that may be known to those skilled in the art are not described in detail herein.

Figure 1 is an illustration of the windowing of a time domain representation of the energy (i.e. excitation) of a coded voiced speech signal. As illustrated, the voiced speech signal may be separated into segments (e.g. windows 101, 102, 103, 104, and 105) before coding. Each segment may contain any number of pitch cycles (i.e. illustrated as big mounds). For instance, segment 101 contains one pitch cycle while segment 104 contains no pitch cycles, and segment 105 contains two pitch cycles. The pitch cycles provide the periodicity of the speech signal.

Periodicity of pitch lag is used in ABS coding approaches such as CELP. One popular approach to detecting the periodicity or pitch lag of a voiced speech signal is the pitch correlation approach. In correlation, one segment of the speech signal is compared to another segment of the signal in order to maximize the correlation between these two segments. The goal is to obtain the pitch lag, which could be small

or large in size, since voiced signal is not purely periodic.

The correlation window is traditionally limited to a certain size in order to obtain a good local pitch lag, a reliable determination of small pitch lags, and an acceptable complexity. However, a problem arises as illustrated in segment 104 where the real pitch lag is larger than the window size and an energy peak is not captured within the target window, which is traditionally on a fixed location.

Since the window size cannot be increased or decreased to cover all potential cases, one or more embodiments of the present invention seeks to maximize the energy in each correlation window by implementing a sliding target window. With this approach, the correlation target window may slide for a known delta in either direction. For example, if the window contains 80 samples, this 80-sample size is maintained, and the location of the target window is allowed to slide by a delta of 20 samples, for example, in either direction thus shifting a range of -20 to +20. The window size remains fixed.

Figure 2 is an illustration of the sliding target window concept in accordance with an embodiment of the present invention. In this illustration, the original window 104 does not capture any peak energy; however, if the correlation window slides to the right by an amount Δt (e.g. N samples), more and more portions of the peak energy 220 is captured within the window (illustrated as window 204). (Note that the slide illustrated in Figure 2 is exaggerated for clarity. In actual implementation, all that is required is to slide the window enough to capture the entirety of peak energy 220). As a result, a better correlation can be achieved between the previous window 103 and the new window 204, while complexity is not affected by maintaining the window size.

This approach is significant for wideband speech processing, since there is

more irregularity or noise in the high frequency areas so that the distance between energy peaks may be more randomly spaced.

It should be noted that the sliding window's computational complexity is minimal since as the window slides, a sample at one end is removed while a new sample at the other end is added to maintain the window size. Therefore, the energy calculations within the sliding window are made without affecting system complexity. Figure 3 is a flowchart illustration of a positive sliding window in accordance with an embodiment of the present invention. Note that the correlation window may slide in either direction (positive or negative).

As illustrated, the total energy **E** within a correlation window of size **N** is computed in block 302. The total energy is the sum of all the energy values, **e**, at each sampling point, **i**, within the correlation window. In block 304 a counter (or sliding index) **j** for the slide width of the sliding window is initialized to zero and the total energy in the current (i.e. initial) window is saved into $\mathbf{E_P}$ in block 306. Also, the current sliding index **j** is saved in $\mathbf{j_p}$. The sliding index counter **j** is incremented in block 308 to move the correlation window to the right. In block 310, a determination is made to assure the maximum delta window shift value is not exceeded. If the maximum slide width is reached, in either direction, pitch correlation is computed by searching for possible pitch lags from the current determined target window and the window at a distant pitch lag.

If, on the other hand, a determination is made in block 310 that the slide width maximum has not been exceeded, a new energy value is computed for the for the new window in block 312 by adding the $\mathbf{(N+j)^{th}}$ energy value to and subtracting the $\mathbf{j^{th}}$ energy value from the total energy **E**. Note that the entire energy is not recomputed. In block 314, a determination is made if a maximum energy value has been found by

checking the newly computed total energy value **E** against the saved energy value $\mathbf{E_P}$. If **E** is greater than $\mathbf{E_P}$, then $\mathbf{E_p}$ and $\mathbf{j_p}$ ($\mathbf{j_p}$ memorizes the best window location) are updated. The computation continues the sliding window process by returning back to block 306 until reaching the maximum shift delta.

If, on the other hand, a determination is made in block 314 that **E** is not greater than $\mathbf{E_P}$, then the computation continues the sliding window process by returning back to block 308 to increment the sliding index counter, **j**, until the maximum shift delta is reached. In block 318, pitch correlation is computed using pitch lag from the current determined target window and the window at a distant pitch lag.

Embodiments of the present invention may slide the window first to the one side, then to the other side in search of the maximum peak energy value. For instance, to move the window to the left may involve simply modifying the equation in block 312 to ($\mathbf{E = E - e_{N-j} + e_{-j}}$), for example, in order to achieve a left shift. The idea is to maximize the energy of the windowed signal by providing at least one peak energy cycle within the correlation window.

Although the above embodiments of the present application are described with reference to wideband speech signals, the present invention is equally applicable to narrowband speech signals.

The methods and systems presented above may reside in software, hardware, or firmware on the device, which can be implemented on a microprocessor, digital signal processor, application specific IC, or field programmable gate array ("FPGA"), or any combination thereof, without departing from the spirit of the invention. Furthermore, the present invention may be embodied in other specific forms without departing from its spirit or essential characteristics. The described embodiments are to be considered in all respects only as illustrative and not restrictive.